Project Overview

The purpose of the Facility Registry System (FRS) project is to provide EPA with a central database of facility identification records that links all facility oriented program system records. It is the long term goal to have the FRS replace the need for EPA programs to separately and individually collect facility identification data for each media national system.

The FRS will be managed by the Office of Information Collection, Office of Environmental Information. OEI, program system managers, regional data stewards and states engaged in data stewardship will provide support in the verification and data ownership creating a central, authoritative record for facility information. Web-based access will provide the reporting community with a mechanism for verification and correction of facility identification data in the FRS.

The FRS project is a primary component of the Information Integration Initiative (I-3) and will be designed to integrate regulatory and place-based information. The I-3 is an effort by the Agency, in partnership with States and Tribes to create a national environmental information exchange network.

Background

Development of the FRS continues the objective of giving environmental managers and the public convenient access to a wide range of information necessary for effective risk-based decision making and multi-media analysis. Traditionally, EPA data systems have been developed by program offices to support the management of separate environmental programs authorized by environmental laws. Information was rarely transferable from one system to another without a great deal of custom intervention. This made management of information regarding regulated facilities difficult to integrate across environmental programs.

In 1990, EPA issued an Agency Order to establish a data standard for unique facility identification codes on facilities regulated by the Agency. The purpose of the Order was to establish a data management standard that would improve the compatibility of all Agency facility data by providing a fundamental piece of "linkable" information for all facility oriented data-collections and records. The Agency needed to establish a standard and a means to integrate records because media programs often labeled the same regulated facility differently. As a result, it was very difficult to determine whether data retrieved from different information systems were actually about the same facility.   The Agency needed a mechanism to tie these disparate records together and optimize the use of the wealth of environmental data collected and stored about facilities.

The Office of Information Resources Management adopted an implementation plan to achieve the

Agency data integration goals. The plan included the management of the Facility INDex System (FINDS), a central database in which each regulated facility would be assigned a unique identifier and no facility be assigned more than one identifier.

In 1998, FINDS was replaced with a web accessible linkage application, the Facility Linkage Application (FLA) using the facility tables in the Agency's Envirofacts Warehouse. The impetus behind FLA was to improve the Agency's ability to accurately link facility identification information in the program systems to each other and to begin to establish an infrastructure to better manage facility identification information.   The FLA is a tool that links facility information from multiple program system records and provides a unique linking number to these records. The Agency established Regional Data Stewards to augment the FLA with manual intervention for records pertaining to facilities that do no link through the FLA.

Migration from FINDS to FLA provided incremental progress. However, EPA is responding to increased need for multi-media information analysis and improved data quality by now building a Facility Registry System of authoritative information and links to media-specific system records.

Project Tasks

Business Requirements for Quality Facility Identification Information in the FRS

The FRS will take advantage of advances in relational database technologies and the significant new information resources available via the world wide web. The FRS will combine good data management processes including automated matching protocols with manual intervention and close partnerships with Regions and States largely responsible for collection of data in EPA national systems and State management systems. The FRS requirements are to:

❏ Provide an authoritative central source of core, accurate information on regulated facilities.
❏ Provide the public with more accurate, more comprehensive search and easier to understand facility identification information.
❏ Improve Agency's ability to do tabular analysis on facilities EPA regulates.
❏ Improve data quality of facility information by eliminating duplicate records, reducing the potential for creating erroneous facility records and filling data gaps.
❏ Reduce burden on regulated community by not having to report identical or similar information on multiple environmental forms.
❏ Assist in the registration and validation of electronic certificates needed for Central Receiving electronic reporting.
❏ Provide regulated entities with an easy way to view the information the Agency maintains and to correct erroneous and out-of-date information.
❏ Allow States and EPA to share integrated facility data.
❏ Allow for easy application development of tools requiring facility data from EPA because

of the documented database structure, model and physical record.

1.   Points of Coordination and Management

The following tasks will be coordinated and managed by the OEI Project Team in coordination with three Action Teams consisting of representatives from program offices (including I-3 Detailees), Regions and states. The three Action Teams are:

❏   Data Management Change Protocols
❏   Data Stewards Roles and Responsibilities
❏   State Data Exchange and Data Quality

2.   Development and Build out of the Data Model and Web Interface

Under the Reinventing Environmental Information Initiative, the Agency developed an initial version of a small facility registry system from the program records of the Toxic Release Inventory (TRI) of 1997 and the Risk Management Plan submissions of 1999. In 2000, (FRS) will build out the data model to be in full compliance with the Agency's facility identification data standard and the Facility Identification Template for States - 2 (FITS-2). This will require that the data model be expanded to include industrial classification codes (SIC and NAICS) and corporate contact, locational data and affiliation information and links to other national systems.

Further refinement of the data model will be necessary to support the registration activities of the Central Receiving (CR) project. CR will rely on the FRS to validate submitter data for electronic signature. The FRS will also provide CR with values to pre-populate forms for individuals at the facilities submitting compliance reports to EPA. Conversely, CR will function as another source of facility identification data through its electronic certificate process to update and keep current FRS values. By incorporating FRS into the CR business process, the concept of central management of facility identification data fulfills a business need..

All enhancement and expansions to the data model will be supported by the development of appropriate metadata to define the structures and documentation of procedures that are to be added to the database.

Public access to authoritative facility information is a primary function of the facility project. Development and maintenance of a well constructed and user friendly web interface is mandatory to support public access. Web access will also be designed to ensure that the regulated community and public can provide feedback and comments on facility records, creating a process for validation and notification of gaps and changes in FRS records.

The web interface supporting the linkage application, which is only available from the

EPA intranet will also need to be expanded and modified to accommodate changes in the FRS data model and to allow data stewards access to the complete database.

3.   Creation of FRS Data Management Procedures for the Facility Record

Successful implementation of the FRS will depend on the establishment of sound business processes based on documented business rules. Business rules are defined as statements that define or constrain some aspect of a business.   From the perspective of the information system, business rules express specific constraints on the creation, updating and removal of persistent data in the information system as it supports some aspect of Agency business. Recognizing that business rule statements indicate discrete operational practices or policies, this project plan has been designed to ensure that the business rules, or operational practices, that govern the FRS are understandable, written using natural language, and allow for sufficient flexibility to make rapid change possible.

The FRS Data Management Procedures will document the validation and verification processes that will be applied to each facility record prior to the record being created in the FRS. The document will lay out the defined set of procedures that locational, environmental interest, organization, contact, and industrial classification data will undergo prior to acceptance into the database. These procedures will include checking for completeness, consistency, and validity and will describe the vigorous automated and manual checking that will be performed to prevent duplicate facility records. The document will also describe the standardization that will be applied to the facility site name and address data elements to aid in automated matching processes and the role of standard reference tables in checking for data quality. Records which do not pass the validation checks will be flagged for manual review and research.

The FRS Data Management Procedures will describe the rules pertaining to derived data values.   A derived data value is created by an inference or a mathematical calculation from other values. The FRS will take advantage of the many new information resources available and rely on derived data where it is appropriate.

The FRS Data Management Procedures will lay out the rules governing updates or changes to FRS records. A series of procedures for changing records based on use or receipt of authoritative data will be established and implemented. Closely linked to the validation and verification procedures, update procedures will vary by data element and data source and reflect the unique conditions in program systems. These procedures will also recognize that update protocols will be different depending on whether a facility record has a one to one relationship with a program record - i.e., only one program (state data system) has an interest in the facility or in multiple agency programs (state systems) have an interest.

The FRS Data Management Procedures will specify the conditions under which facility records that exist in more than one program system can be automatically linked and when manual

intervention will be required. The Procedures will specify the comparison that records will undergo and the formulas that will govern the linkage process, how and when new linkages are created or broken and the specific conditions that will trigger manual intervention by the Data Steward(s).

The Records Change Protocol Action Team has begun developing update rules by reviewing each data element and soliciting program system views on appropriate management requirements. This Data Management Action Team will also consider which data elements are the minimum required for creation of a record in the FRS and the supporting business rules.

4.   Population of the FRS Database

Initial Population Plan

Facility/site data will be obtained from a variety of sources including EPA program systems, Central Receiving registration, State master records and from outside parties. Current records suggest that the Agency has an active interest in approximately 750,000 facilities. Of the 750,000, approximately 50,000 - 60,000 are considered "major" as defined by the various program definitions. The planned initial population sequence is:

FY 2000
> Populate data elements from RMP-99 and TRI-98
> Populate and validate with Central Receiving registered facilities when not included through program system records
> Populate with RCRA TSDs - involved in OSWER-GPRA plans (6200 facilities)
> Populate with PCS Major facilities (4500 facilities)
> Populate with State Master Records(number to be determined by funding, timing and the number of state systems that volunteer for data sharing)
> Populate with TRI 1997 universe of reporters (2,000 facilities)
> Populate with 1995 & 1997 BRS reporters (20,500 minus most TSDs)

FY 2001
> Populate with AIRS facility majors (90,000 facilities)
> Populate with orphaned TRI and RMP facility records that were incomplete data (250 records).
> Populate with the remaining Large Quantity Generators in RCRIS (few 1000s)
> Populate with SDWIS Majors where possible (20,000 entities)
> Populate with any of the SFIP facilities still not picked up in the above.
> Continue to populate with State Master Records(number to be determined by funding, timing and the number of state systems that volunteer for data sharing)

This should bring the universe of facilities into FRS to above 250,000 unique entries, covering the universe of "majors" and state master records.

Discussion with Program Systems

The creation of the FRS record starting with program records involves support, cooperation and interaction with the program system national managers. The population of the FRS record will require the assistance from the program offices in the following tasks:

S   identification of the facility data elements in the national system that contain the facility identification data that will be used to populate the FRS records.

S   a mapping of those elements in the national systems to the FRS data model

S   establishment of regular update or refresh cycles (or extracts of the facility data values from national program systems) with the FRS database. Regular access or extract routines need to be scheduled to ensure a program record is still active or part of the national program. Program records that are no longer part of the national system need to be archived in the FRS data. Timely updates or extract routines will be negotiated on a system by system basis with the national system manager.

S   regular interaction with the FRS team after data reconciliation takes place in the FRS creation and linkage processes. Program records with inadequate data or inconsistent data will have to be manually reconciled first at the national system level and, where needed, at the Regional and state level. This interaction will be required to ensure all program records have an FRS record and that all program system Ids are accounted for in the FRS.

S   Program Participants In FRS Action Teams: Charles Spooner, OW (I-3); Jeff Sexton, OW; John Sullivan, OEI; Dave Levy, OSWER/OSW; James Rothwell, OIG; Bewanda Alexander, OPPTS Ron Decesare, OW (I-3); Eugene Durman, OAR (I-3); Leonard Fitch, OW; Lisa Jenkins OSWER (I-3); Merle Miller, OECA; Dina Villari, OSW; Jeff Sexton, OW; Laura Milton, OECA; Allen Abramson, OPPTS; David Meisenheimer (OAR): Abe Segal, OW

Data Challenges with the Program Systems

The program system universe of regulated entities will present major challenges to the FRS database and the Facility Identification standards. Facilities are deemed places of the earth. However, many program system records are for transporters, mobile units, items at sea, individuals and groups of individuals. In the program systems, many places can be associated with one unique regulatory activity or permit. Issues of contractor operated on government lands, government-operated and contractor-owned, and operations on Indian lands or foreign countries, can present FRS with unique challenges on having authoritative records pertaining to a facility. As a part of the I-3, this will be a part of efforts by the Agency, in partnership with States and Tribes to create a national environmental information exchange network.

The FRS project is a cornerstone for the Information Integration Initiative (I-3) and should be designed to integrate regulatory and place-based information with facility identification information. The name of FRS as a "facility" system will be difficult to work through as

regulatory and placed based information needs are identified for objects such as monitoring stations and entities that have enforcement actions taken against them. This minor challenge will be worked through as activities of cooperation are identified and resolved.

5.   Data Stewards Network Organization

The FRS will be supported by a network of data stewards, as well as, the Headquarter system managers. The data stewards will act as "quality champions" for quality facility data. They will guide the establishment of accurate environmental linkages between records from the program systems based on their knowledge of the places in the regions and state. The roles and responsibilities of the data stewards are being finalized by one of the above workgroups, and will include roles such as:

❏   Acting as the chief point-of-contact for the FRS development team.
❏   Acting as a liaison between Regional program system data coordinators and State Data Stewards on facility identification data.
❏   Coordinate training activities for system users.
❏   Reviewing facility data to determine which discrepancy or feedback reports require action and which facility records require manual resolution.
❏   Resolving data quality issues and program coordination which require human intervention.
❏   Assisting in identifying enhancements for the FRS application.

On-going Data Stewards Action Team meetings will result in the production of a Data Stewards Procedures Manual that can be adopted by all regions with sufficient flexibility to account for unique relationships between states and regions. This effort should be completed by Spring.

A presentation of these roles and responsibilities to regional personnel is planned. It will explain both the FRS and the Data Stewards responsibilities. Initial meetings should be started by mid-April.

A national meeting for Data Stewards will be organized for May 2000. This meeting will provide an opportunity for regional representatives to help develop basic stewardship training material, learn more of the duties expected of a data steward, hear from OEI representatives and meet with national program managers.

6.   Regional Data Cleanup Priorities

Regions will establish a 2year strategy to improve the quality of facility-based information. This 2-year strategy is linked directly to expectations for continued funding support. The strategy will be developed based on thorough review of existing records and knowledge of regional media programs and participating States.   Although regional priorities will be tailored to the specific

problems in a region, there are general activities that all regions should plan for:
- primary attention to the FRS records that are incomplete.
- primary attention to manual linkages of program records to the FRS in the population strategy.
- maintaining a network of regional and state contacts for data stewardship.
- primary attention to discrepancies and errors identified by the public or regulated community pertaining to the FRS records.
- working with selected states when master state record files are being reconciled in the FRS.

Each Region will have unique issues and activities from states or the region that will be need to be addressed at a priority level to achieve success. These activities will need to be identified and discussed with appropriate OIC management. Possible regional areas of priority for long term planning might include:

Year One Priorities:

1-1. Priority attention will be focused on the new facility records in the FRS. As the data population activities are achieved, the Primary Data Stewards with the assistance from the Media Data Stewards, will review and complete the FRS records with data from various regional and state resources. Ensuring that the FRS record is complete with verified values will be a high priority as the facility records are created. Ensuring that all program records associated with the FRS record are identified is important and a high priority.

1-2. For Primary Data Stewards: Work with Media Data Stewards to improve worst quality data elements. Begin work on obvious linkages problems -i.e. those not yet linked and those mis-linked.

1-3. For Media Data Stewards : Identify and begin work on the three worst data quality elements within each media program identified to the FRS

Year Two Priorities:

2-1. Priority attention will be focused on the new facility records in the FRS. As the data population activities progress, the Primary Data Stewards with the assistance from the Media Data Stewards, will review and complete new FRS records with data from various regional and state resources. Ensuring that the new and existing FRS records are complete and current with verified values will be a high priority as the facility records are created and maintained. Ensuring that all program records associated with the FRS record are identified continues to be an important and high priority.

2-2. For Primary Data Stewards: Continue work with media program Data Stewards and

further work on linkage improvements.

2-3. For Media Data Stewards: Continue cleanup of worst quality data elements and begin linkage improvements especially those involving consistency in naming convention and addresses between and among programs tracking the same facility.

2-4.   Work with State Master records will be a larger priority. Working with the State and Regional Media Data Stewards will be important to ensure that all state and EPA records are complete, quality assured and properly linked.

Priority Setting Based on Quality Diagnostics

In reviewing Regional data we find, primarily, two scenarios impacting data quality. The first is program records with such poor quality information that it is impossible to link the record or identify if the record is associated with other program records. The second is linked records with inconsistencies: possible missing information, inaccurate information, or lack of convention in facility naming. In these instances, the records are typically found to be accurately linked, but contain poorly constructed individual program records.

It is the intent to have the FRS record, which is populated not just from State and program data values, but from other sources that provide correct values, to ensure data quality is high and linkages are verified.

The single largest problem with program data, and more likely to occur in rural areas, is use of the mailing/corporate address instead of the physical location. Where physical location is used, the various data systems handle the identification differently.

We expect the Regions to continue data improvement work utilizing tools available. As we transition into the Facility Registry System we see more opportunities to contribute high quality data linkages to the agency's integration efforts.

7.   State Master Record Utilization

The decisions regarding utilization of state master records will be considered in a State Data Exchange and Data Quality Action Team including representatives from state environmental information programs. This Action Team will consider:

•   how to utilize state master records as the FRS database is populated;

•   how to exchange data on a regular and efficient basis; and determine what verification and stewardship processes will be necessary to insure consistent data quality for all records from both EPA and state sources.

        This Action Team is expected to be operational by April and provide recommendations in May of 2000. Specific discussions of data exchange issues will be held in the context of needs for Central Receiving and general approaches for data exchanges networks between EPA and states.

        It is expected that data sharing with the States will involve a two way exchange. It is expected that data will be shared to EPA and also back to the State. As can be expected with large data file transfers, extensive discussions will be required to ensure a smooth data exchange takes place. Further discussions will focus on what data to share. State environmental interests will be of particular concern, since that item will greatly effect the universe size and linkage processing to the facility record in FRS and the federal program interests in the FRS. It is expected that EPA will not be interested in all state records. Individual state discussions will provide the specifics in what data is transferred. Discussions will primarily focus on the data elements that are part of the FITS-2 data model.

        It is expected that States will have a strong interest in the FRS records that have program linkages to the Risk Management Plan (CAA 112), to the Toxic Release Inventory program, the Superfund program, the PCB handlers, and the EPA Docket system. The FRS linkages to the Permit Compliance System, RCRIS hazardous waste handlers, the Biennial Reporting System of hazardous waste handlers, AIRS data, Safe Drinking Water, and other federal delegated programs should be of value to verify state master records and program linkages.

8.   Outreach Activities

•   Regional Offices - These efforts include Action Team participation, FRS operation and Data Stewardship briefings for regional staff. It is planned that Regional / State meetings organized by the Regions will provide for Regional and State interaction and communication that will further our understand of the FRS relationships with state and regional data and systems.

•   FRS Data Stewardship Action Team Regional Participants:

        Ken Blumberg, R-I; Mae Dooley & Joseph Kunz, R-III; Steven Goranson, R-V; Sam Ferrel, R-IX; Fred Lief, David Tetta & Robert Fallis, R-X; Rebecca Kemp & Rich Nawyn, R-IV; Robert Fallis, R-X; Joe Kunz, R-III; Noel Kohl, R-V; Cynthia Culvert ABT Associates;

•   States - Outreach will include Action Team participation, Regional activities and appropriate briefings for the IMWG.

•   FRS State Action Team Participants

> Cathy Wagenfer, MD; Edward Meyer, MN; Lynn Singleton, WA; Ron Tuminski, Irene Kropp and Betty Miller,NJ; Kimberly Nelson, PA;Zenus Brehm, PA:

- Industrial/Reporting Facilities - Specific briefings for industry coalitions and trade groups will be scheduled. Contacts to initiate this effort have been made.

- Environmental Special Interest Groups - Contacts with environmental groups will be initiated to inform them of the process and progress.

9.   Training

Training will be developed to deal with items related to Stewardship roles and responsibilities especially creation/deletion of linkages in the linkage application component of the FRS. Training will be computer based to the extent possible to limit the need for travel for state and regional personnel.

10.   Project Milestones

1.   Points of Coordination and Management

3/1	Establish Action Teams
3/1 - 4/20	Action Teams Meet and Develop Products
4/20	Action Team Products Delivered

2.   Development and Build out of Data Model and Web Interface

3/1 - 9/30	Data Model expanded
4th qrtr	FRS web site goes live

3.   Population Schedule

Present through 5/15	Build out of FRS with TRI & RMP Data
5/15 - 5/30	RCRA TSDs, PCS Majors,
6/1 - 8/30	BRS, state master records
by 9/30	Central receiving facilities and State master records
thru 9/30	Negotiations with program systems

4.   Creation of FRS Data management Procedures

Through 9/30	Refine verification procedures and data quality regimens
By 6/30	Document Records Change Protocols
9/30	Publish FRS Data Management Procedures on FRS web site

5.   Regional Data Stewards Network Organization
        4/20   Produce Draft Data Stewards Manual
        4/15 - 8/15   Presentations on Stewardship roles and responsibilities
        May   National Meeting of Data Stewards
        9/30   Publish final Data Stewards Manual on FRS web site

6.   Regional Data Cleanup Priorities
        5/30   Develop Regional Strategy
        6/30 ongoing   Implement Strategy

7.   State Master Record Utilization
        Present thru 5/1   Action Team Meets
        Early May   Action Team Recommendations delivered
        Present thru 9/30   Negotiations with States regarding population with State
                                        Master Records

8.   Outreach Activities
        4/1 - 5/30   Develop Briefing schedule
        4/30 - 9/30   Give briefings

9.   Training
        7/1 - 9/30   Initial Training made available via the web

11.   FRS Budget (000s)

        FRS-1 - Database Population

The FRS project will create over 50 thousand authoritative records on facility identification from
subsets of the eight national program system. This will also include the transfer of data from
several state master records.

        FRS-2 - Regional Data Steward Support

Regional Data Steward Support - Will provide support for regional efforts in each of the 10
regional offices to support record verification, research and linkages for the facility registry
systems records. This effort will also provide liaison to states. This will also provide partial
support for development of training module for the data stewards.

        FRS-3 - Database Model Development

This process will develop the full data model for the facility registry database. This includes all
the major data groups contained in the facility information template for the states (FITS/FITS2).

This contains at a minimum the Facility Identification Standard data elements. This effort will insure integration with Central Receiving.

       FRS-4 - Database M&O

This will provide contractual and other support for a process of regular data refresh from the national systems used to create and verify the facility records. This process will utilize data from national system collections, states master records, central receiving.

       FRS-5 - Develop Web Interface

Development of Web Interface - Provide contractual support to develop web based searches, retrievals and reporting tools for the Internet. This will be used for both internal EPA and states as well as for public access. This will include an interface for reporting facilities to verify FRS records. This will also include development of web-based training module for data stewards.